DOCUMENT RESUME

ED 320 946 TM 015 216

AUTHOR Des Marchais, Jacques E.; And Others
TITLE Theory-Guided Design of a Rating Scale for Program Evaluation in Problem-Based Medical Curricula.
PUB DATE Apr 89
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989). For a related document, see TM 015 215.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Construct Validity; *Curriculum Evaluation; Factor Analysis; Foreign Countries; Higher Education; Medical Schools; *Medical Students; *Program Evaluation; *Rating Scales; *Research Design; *Test Construction; Test Reliability; Test Validity; Theory Practice Relationship
IDENTIFIERS *Problem Based Medical Curriculum; Quebec

ABSTRACT

The design and evaluation of a rating scale for program evaluation in problem-based medical curricula are described. The design process was guided by a theory of how students learn in problem-based learning as modeled by W. H. Gijselaers and H. G. Schmidt (1988). The rating scale was used in an evaluation of a new problem-based, community-oriented curriculum at the University of Sherbrooke in Quebec (Canada). A 46-item rating scale was constructed using problem-based learning theory. After each unit of the curriculum, the rating scale was administered to the 95 students in the first year of the new curriculum. Construct validity was tested by factor analysis. The internal consistency of the scale and interrater reliability were assessed. Criterion validity was investigated through a regression analysis. The 10 subscales of the rating scale appeared to cover the most relevant constructs of the problem-based approach. Subscales generally demonstrated sufficient internal consistency and interrater agreement, but criterion validity was rather low, which is attributed to the way student achievement was measured. This theory-guided approach to program evaluation contrasts sharply with procedures emphasizing "bottom-up" construction of a rating scale. Five tables present study data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320946

Theory-Guided Design of a Rating Scale for Program Evaluation in Problem-Based Medical Curricula

Jacques E. Des Marchais, Henk G. Schmidt° and Robert Black
Université de Sherbrooke, Québec, Canada and
°University of Limburg, The Netherlands

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

HENK G. SCHMIDT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Meeting of the American Educational Research Association, San Francisco, CA., 1989

TM015216

BEST COPY AVAILABLE

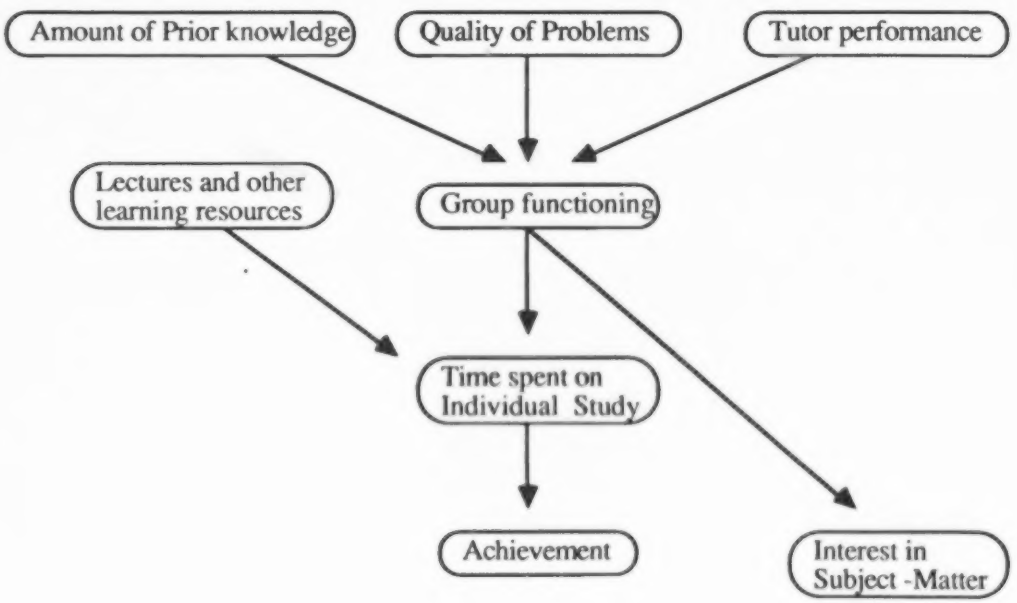
Program evaluation constitutes an important instrument for quality control in higher education. Its outcomes may influence the content and design of courses, tenure perspectives of teachers, or even accreditation decisions by professional bodies (Conrad & Blackburn, 1985). Student ratings of the quality of instruction constitute an easily accessible source of informa-tion in this respect, particularly, since they usually can be shown to have sufficient reliability and validity (Cohen, 1981).

In this paper, the design and evaluation of a rating scale for program evaluation in problem-based medical curricula will be described. The way in which the rating scale was constructed however, deviates from other approaches common to the field (e.g. Marsh, 1982) in that a theory on how students learn in a problem-based program and how this learning is affected by instructional procedures was used to guide the the design process. Usually, an empirical approach is followed in which the selection of items is based on the designer's intuitive notions about what constitute important elements of the instructional process. Such a "bottom-up" procedure takes the risk of producing a rating scale that does not so much reflect the important elements of the actual learning situation as measuring mere idiosyncrasies about education which appear to be salient simply because they are shared by students and designer alike. A theory-guided or "top-down" approach avoids this kind of confusion.

Ingredients for a theory of problem-based learning. Problem-based learning can be characterized as follows: A collection of carefully constructed problems is presented to small groups of students. In medical education, they usually take the form of a description of a patient, presenting with a complaint and with a number of signs and symptoms. The task of the group is to discuss these problems and produce tentative explanations for the phenomena described in terms of some underlying process, principle or mechanism. Essential to the method is that students' prior knowledge of the problem is, in itself, insufficient to understand it in depth, so that during initial analysis dilemma's will arise and questions will come up that can be used as learning goals for subsequent individual self-directed learning. While analyzing a problem in a prescribed, systematic fashion, the group is guided by a tutor. His or her task is to stimulate the discussion whenever necessary, to provide students with subject-matter information when adequate and to evaluate progress being made. References, audiovisual aids and occasional lectures are included as learning resources relevant to the understanding of the problems.

This approach to instruction has been modelled by Gijselaers and Schmidt (1988). They have formulated their theory within the context of the "models of school learning" tradition represented by authors like Bloom (1976), Carroll (1963) and Cooley and Leinhardt (1980). Figure 1, represents the Gijselaers and Schmidt model in a somewhat simplified manner. The model can be considered a causal and quantitative representation of the learning going on within a problem-based context. According to the model, an increase in the magnitude of one of the variables characteristically causes an increase of the magnitudes of other variables. The arrows indicate the direction of the causal influence. For instance, this theory predicts that an improvement in the quality of the problems presented to students, all other things being equal, will result in improved group functioning. Better executed small-group tutorials, in turn, affect study time, which leads to higher achievement of the students involved. The role of the other variables involved can be interpreted in much the same way.

Figure 1. Causal model of problem-based learning adapted from Gijselaers and Schmidt (1988)



In a study involving more than five hundred subjects in the University of Limburg's problem-based medical curriculum, Gijselaers and Schmidt (1988) tested the model exemplified in Figure 1 using a path-analytic approach. During one academic year, they monitored the learning of 240 tutorial groups in 20 six-week courses or 'units'. Results on the various end-of-unit examinations were used as a measure of academic achievement. All other variables were operationalized by sets of descriptive items rated by students and tutors. Interrater agreements and internal consistencies of the various sets were high and students' judgments were shown to have considerable construct validity (Gijselaers, 1988). Generally, their results corroborated the model's specifications. For instance, they found path coefficients of .43 between quality of problems and group functioning, .32 between tutor behavior and group functioning, .42 between time spent onindividual learning and achievement and .57 between group functioning and interest in subject-matter. The causal relationship between group functioning and time spent however, was negative --expressed in a path coefficient of -.24 -- which the authors explained as a pay-off effect: The better a tutorial group functions (the harder it works, the more ground it covers, etc.), the less time its members need for individual study.

In summary, this study shows that variables relevant to the instructional process generally, entertain a unidirectional, causal, relationship with student achievement. An implication of these findings is that achievement can be improved by improving the quality of instruction.

METHOD

Subjects and Curriculum. Subjects were 95 first-year students of the Faculté de Médecine of the Université de Sherbrooke, Québec, Canada. This medical school is presently involved in a large-scale transition process in which a four-year discipline-based conventional curriculum is gradually being changed to a problem-based, community-oriented program. Therefore, careful program monitoring was considered an essential element of the innovation. The first group of students enrolling in the renewed curriculum started September 1987 and has by now completed seven six-week courses or "units" which comprise the first curriculum year.

Design of the rating scale. Based on the theoretical notions concerning problem-based learning outlined above, a 46-item rating scale was constructed. For each of the constructs mentioned in Figure 1, a set of items was written that covered various facets of the variable concerned. Most of the items consisted of a statement and a 5-point Likert-scale, ranging from "highly disagree" to "highly agree", to which the students could respond by encircling a number (See Table 1 for examples of items).

Table 1. Sample items for each of the variables in the Gijselaers and Schmidt (1988) model

Prior knowledge/Difficulty level :
The unit's subject-matter was adapted to my prior knowledge
The unit's subject-matter was difficult to understand
Quality of problems :
The problems were clearly stated
The problems sufficiently stimulated group discussion
Tutor performance:
The tutor's questions stimulated the discussion
The tutor has given me feedback during formative evaluation
Group functioning:
The meetings stimulated self-directed learning activities
I found the atmosphere in my group agreeable
Learning resources:
The lectures related well to the subjects I studied
The references proposed were adequate
Study time:
How much time on the average did you spent each week on independent study? (Fill in the answer in whole hours)
Interest:
I've learned a lot during this unit
I consider the subject of this unit interesting

Student achievement was measured three times during the year, using a test consisting of open and multiple-choice questions.

Procedure. After each unit, the rating scale was administered to all students. The average response rate was 86%. The total number of rating scales returned was 566. In order to evaluate the measurement characteristics of the instrument, a series of statistical analyses were carried out. First, factor analysis using principal components and oblique rotation[1] was conducted in order to determine to what extent the theoretical constructs were adequately represented in the rating scale. Second, the internal consistency of the resulting scales was checked. Third, since sufficiently high interrater agreement was considered a necessary condition for an instrument used for program evaluation, intraclass coefficients were computed for each item. Fourth, criterion validity was investi-gated by conducting a regression analysis of the instructional variables on achievement.

[1]The dimensions underlying the questionnaire are not considered independent. On the contrary, the theoretical considerations that led

RESULTS and DISCUSSION

Factorial Structure. Table 2 displays the result of the extraction of principal components before rotation. The Kaiser normalization criterion was applied to determine the number of components to be used for further analysis[1].

Table 2. Eigenvalues and percentages of variance explained by components extracted

Factor	Eigenval.	% of Var. Explained	Cumul. % of Var.
1. Tutor as contributor to learning	10.5	24.5	24.5
2. Satisfaction with unit	4.7	9.4	33.9
3. Group functioning	2.7	6.2	40.1
4. Quality of problems	2.6	5.9	46.0
5. Adequacy of lectures	1.9	4.5	50.5
6. Tutor as personal guide	1.5	3.4	53.9
7. Breadth of learning	1.4	3.3	57.2
8. Prior knowledge/Difficulty level	1.3	2.9	60.1
9. Time-on-task	1.1	2.6	62.7
10. Systematic approach	1.0	2.3	65.0

The principal components analysis extracted 10 factors with an eigenvalue greater than 1.0. These factors together explain 65% of the total variance in the data. The number of components extracted is fairly large, if one takes into account the number of items involved (43). In addition, the amount of variance explained by the first component is rather small. These results indicate that the rating scale has a true multidimensional structure and that a Halo effect as a possible explanation for the response pattern of students is unlikely (the first factor would, in that case, explain more of the variance). By comparison: The SEEQ (Students' Evaluations of Educational Quality), a well-known 35-item rating scale, considered an adequate instrument for measuring instructional effectivity, consists of nine factors explaining 88% of the variance. However, its first factor explains no less than 56.8% of the total variance and only five factors have a eigenvalue greater than 1 (Marsh, 1982).

The ten factors resulting from the analysis are easily interpreted, considering the contents of the relevant items. In addition, they appear to cover quite well the theoretical constructs discussed in the introduction section. An exception is the "Interest in subject-matter" construct, that emerges in other studies as one of the outcomes of the learning. It is not clear why this is the case. Interest as measured by the item: "I consider the subject of this unit interesting" does not load sufficiently on any of the factors included. However, another affective dimension, usually not mentioned in relevant studies, described here as "General satisfaction with the course", prominently emerges from the data analysis.

Factor loadings of the items comprising the dimensions uncovered in the present study, range between .84 and .44.

Reliability analyses. Table 5 contains results of the reliability analyses. Although the number of items

to its development presupposes certain relationships among variables.

[1]The so-called "Scree"-test is sometimes used for deciding on the number of components (Cattel, 1966). This test however, typically leads to the inclusion of no more than three (in rare cases four) factors in the analysis irrespective of the nature of the data. Since the present questionnaire was based on theoretical considerations, it was important to check to what extent the relevant constructs have been operationalized in the items included, independent of the issue of how much variance each construct explains.

constituting the

Table 5. Internal consistencies and average interrater agreements

	N items	Alpha	Interrater agreement
1. Tutor as contributor to learning	8	.82	.85
2. Satisfaction with unit	5	.81	.89
3. Group functioning	6	.92	.87
4. Quality of problems	5	.88	.76
5. Adequacy of lectures	4	.83	.87
6. Tutor as personal guide	4	.88	.87
7. Breadth of learning	2	.71	.69
8. Prior knowledge/Difficulty level	3	.73	.90
9. Time-on-task	1		.63
10. Systematic approach	1		.83

various scales is rather small, the internal consistency of the various scales is high. This result suggests that each of the scales can be considered to measure various facets of a unidimensional and homogeneous construct.

Interrater agreements were computed by averaging intraclass coefficients on each item over each group and each unit. Subsequently, mean coefficients for each of the scales were computed. Since the intra-class coefficient provides an unbiased estimate of the reliability of individual and group judgements, the data summarized in Table 5 indicate that, generally, different judges appear to agree on ratings given to different aspects of the learning situation.

Validity. In order to acquire an estimate of the criterion validity of the rating scale, a multiple regression of factor scores on student achievement was conducted. Multiple R was .53, indicating that the ten variables explained only about 30% of the variance in student achievement. Since student achievement was not measured at the end of each unit but only after two or three units and thus, rating of units had to be combined, this may account for the rather low criterion validity. Construct validity analyses are in progress and will be presented at the time of the conference.

CONCLUSIONS

The development of a rating scale for program evaluation in problem-based medical curricula described in this paper, documents that it is possible to incorporate theoretical notions concerning learning and instruction into an instrument designed to provide data for curriculum reform. The approach followed here sharply contrasts with existing practices which emphasize "bottom-up" construction of rating scales.

The ten sub-scales resulting from principal components analysis of responses provided by first-year medical students of the Université de Sherbrooke, rating aspects of seven six-week units, appear to cover most relevant constructs proposed in Gijselaers and Schmidt's (1988) theory of problem-based learning. An exception was the construct of "interest in subject-matter". It was concluded that in comparison to other instruments available for course evaluation, like the SEEQ (Marsh, 1982), the measurement characteristics of the present rating scale appear to be more than adequate. The subscales generally demonstrated sufficient internal consistency and interrater agreement. In addition, the data showed few signs of being the result of a Halo effect. Criterion validity however, turned out to be rather low. This was attributed to the way student achievement was measured in the present study.

REFERENCES

Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York, NY: McGraw-Hill

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Cattel, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*, 281-309.

Conrad, C. F., & Blackburn, R. T. (1985). Program quality in higher education. In J. C. Smart (Ed.). *Higher education: Handbook of theory and research*. New York, NY: Agathon Press.

Cooley, W. W., & Leinhardt, G. (1980). The instructional dimensions study. *Educational Evaluation and Policy Analysis*, 7-25.

Gijselaers, W. H., & Schmidt, H. G. (1988). The development and evaluation of a causal model of problem-based learning. In Z. Nooman, H. G. Schmidt, & E. Ezzat (Eds.). *Innovation in Medical Education: An Evaluation of its Present Status*. New York, NY: Springer Publishing.

Marsh, H. W. (1982). SEEQ: A reliable, valid and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*, 77-95.